

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: HOPWOOD, JOHN JOSEPH; SCOTT, HAMISH STEELE; WEBER, BIRGIT; BLANCH, LIANNE; ANSON, DONALD STEWART

(ii) TITLE OF INVENTION: SYNTHETIC MAMMALIAN  
α-N-ACETYLGLUCOSAMINIDASE AND GENETIC SEQUENCES ENCODING SAME

(iii) NUMBER OF SEQUENCES: 6

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: NIXON PEABODY LLP  
(B) STREET: 990 STEWART AVENUE  
(C) CITY: GARDEN CITY  
(D) STATE: NEW YORK  
(E) COUNTRY: UNITED STATES  
(F) ZIP: 11530

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk  
(B) COMPUTER: IBM PC compatible  
(C) OPERATING SYSTEM: PC-DOS/MS-DOS  
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: 09/077,354  
(B) FILING DATE: 22-APRIL-1999

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: PCT/US96/00747  
(B) FILING DATE: 22-NOV-1996

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: POKALSKY, ANN R.  
(B) REGISTRATION NUMBER: 34,697  
(C) REFERENCE/DOCKET NUMBER: 2249/104

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 516 742 4343  
(B) TELEFAX: 516 742 4366

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2575 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens  
(F) TISSUE TYPE: Peripheral Blood  
(G) CELL TYPE: Leukocyte

(ix) FEATURE:

(A) NAME/KEY: CDS  
(B) LOCATION: 102..2330

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CCCGGGCTTA GCCTTCGGGT CCACGTGGCC GGAGGCCGGC AGCTGATTGG ACGCGGGCCG 60

CCCCACCCCC TGGCCGTCGC GGGACCCGCA GGACTGAGAC C ATG GAG GCG GTG 113  
Met Glu Ala Val  
1

GCG GTG GCC GCG GCG GTG GGG GTC CTT CTC CTG GCC GGG GCC GGG GGC 161  
Ala Val Ala Ala Ala Val Gly Val Leu Leu Leu Ala Gly Ala Gly Gly  
5 10 15 20

GCG GCA GGC GAC GAG GCC CGG GAG GCG GCG GCC GTG CGG CCG CTC GTG 209  
Ala Ala Gly Asp Glu Ala Arg Glu Ala Ala Val Arg Ala Leu Val  
25 30 35

GCC CGG CTG CTG GGG CCA GGC CCC GCG GCC GAC TTC TCC GTG TCG GTG 257  
Ala Arg Leu Leu Gly Pro Gly Pro Ala Ala Asp Phe Ser Val Ser Val  
40 45 50

GAG CGC GCT CTG GCT GCC AAG CCG GGC TTG GAC ACC TAC AGC CTG GGC 305  
Glu Arg Ala Leu Ala Ala Lys Pro Gly Leu Asp Thr Tyr Ser Leu Gly  
55 60 65

GGC GGC GCG GCG CGC GTG CGG GTG CGC GGC TCC ACG GGC GTG GCG 353  
Gly Gly Ala Ala Arg Val Arg Val Arg Gly Ser Thr Gly Val Ala  
70 75 80

GCC GCC GCG GGG CTG CAC CGC TAC CTG CGC GAC TTC TGT GGC TGC CAC 401  
Ala Ala Ala Gly Leu His Arg Tyr Leu Arg Asp Phe Cys Gly Cys His  
85 90 95 100

GTG GCC TGG TCC GGC TCT CAG CTG CGC CGT CCG CGG CCA CTG CCA GCC 449  
Val Ala Trp Ser Gly Ser Gln Leu Arg Leu Pro Arg Pro Leu Pro Ala  
105 110 115

GTG CCG GGG GAG CTG ACC GAG GCC ACG CCC AAC AGG TAC CGC TAT TAC 497  
Val Pro Gly Glu Leu Thr Glu Ala Thr Pro Asn Arg Tyr Arg Tyr Tyr  
120 125 130

CAG AAT GTG TGC ACG CAA AGC TAC TCC TTC GTG TGG TGG GAC TGG GCC 545  
Gln Asn Val Cys Thr Gln Ser Tyr Ser Phe Val Trp Trp Asp Trp Ala  
135 140 145

CGC TGG GAG CGA GAG ATA GAC TGG ATG GCG CTG AAT GGC ATC AAC CTG 593  
Arg Trp Glu Arg Glu Ile Asp Trp Met Ala Leu Asn Gly Ile Asn Leu  
150 155 160

GCA CTG GCC TGG AGC GGC CAG GAG GCC ATC TGG CAG CGG GTG TAC CTG Ala Leu Ala Trp Ser Gly Gln Glu Ala Ile Trp Gln Arg Val Tyr Leu 165 170 175 180	641
GCC TTG GGC CTG ACC CAG GCA GAG ATC AAT GAG TTC TTT ACT GGT CCT Ala Leu Gly Leu Thr Gln Ala Glu Ile Asn Glu Phe Phe Thr Gly Pro 185 190 195	689
GCC TTC CTG GCC TGG GGG CGA ATG GGC AAC CTG CAC ACC TGG GAT GGC Ala Phe Leu Ala Trp Gly Arg Met Gly Asn Leu His Thr Trp Asp Gly 200 205 210	737
CCC CTG CCC CCC TCC TGG CAC ATC AAG CAG CTT TAC CTG CAG CAC CGG Pro Leu Pro Pro Ser Trp His Ile Lys Gln Leu Tyr Leu Gln His Arg 215 220 225	785
GTC CTG GAC CAG ATG CGC TCC TTC GGC ATG ACC CCA GTG CTG CCT GCA Val Leu Asp Gln Met Arg Ser Phe Gly Met Thr Pro Val Leu Pro Ala 230 235 240	833
TTC GCG GGG CAT GTT CCC GAG GCT GTC ACC AGG GTG TTC CCT CAG GTC Phe Ala Gly His Val Pro Glu Ala Val Thr Arg Val Phe Pro Gln Val 245 250 255 260	881
AAT GTC ACG AAG ATG GGC AGT TGG GGC CAC TTT AAC TGT TCC TAC TCC Asn Val Thr Lys Met Gly Ser Trp Gly His Phe Asn Cys Ser Tyr Ser 265 270 275	929
TGC TCC TTC CTT CTG GCT CCG GAA GAC CCC ATA TTC CCC ATC ATC GGG Cys Ser Phe Leu Leu Ala Pro Glu Asp Pro Ile Phe Pro Ile Ile Gly 280 285 290	977
AGC CTC TTC CTG CGA GAG CTG ATC AAA GAG TTT GGC ACA GAC CAC ATC Ser Leu Phe Leu Arg Glu Leu Ile Lys Glu Phe Gly Thr Asp His Ile 295 300 305	1025
TAT GGG GCC GAC ACT TTC AAT GAG ATG CAG CCA CCT TCC TCA GAG CCC Tyr Gly Ala Asp Thr Phe Asn Glu Met Gln Pro Pro Ser Ser Glu Pro 310 315 320	1073
TCC TAC CTT GCC GCA GCC ACC ACT GCC GTC TAT GAG GCC ATG ACT GCA Ser Tyr Leu Ala Ala Ala Thr Thr Ala Val Tyr Glu Ala Met Thr Ala 325 330 335 340	1121
GTG GAT ACT GAG GCT GTG TGG CTG CTC CAA GGC TGG CTC TTC CAG CAC Val Asp Thr Glu Ala Val Trp Leu Leu Gln Gly Trp Leu Phe Gln His 345 350 355	1169
CAG CCG CAG TTC TGG GGG CCC GCC CAG ATC AGG GCT GTG CTG GGA GCT Gln Pro Gln Phe Trp Gly Pro Ala Gln Ile Arg Ala Val Leu Gly Ala 360 365 370	1217
GTG CCC CGT GGC CGC CTC CTG GTT CTG GAC CTG TTT GCT GAG AGC CAG Val Pro Arg Gly Arg Leu Leu Val Leu Asp Leu Phe Ala Glu Ser Gln 375 380 385	1265

CCT GTG TAT ACC CGC ACT GCC TCC TTC CAG GGC CAG CCC TTC ATC TGG	1313
Pro Val Tyr Thr Arg Thr Ala Ser Phe Gln Gln Pro Phe Ile Trp	
390 395 400	
TGC ATG CTG CAC AAC TTT GGG GGA AAC CAT GGT CTT TTT GGA GCC CTA	1361
Cys Met Leu His Asn Phe Gly Gly Asn His Gly Leu Phe Gly Ala Leu	
405 410 415 420	
GAG GCT GTG AAC GGA GGC CCA GAA GCT GCC CGC CTC TTC CCC AAC TCC	1409
Glu Ala Val Asn Gly Gly Pro Glu Ala Ala Arg Leu Phe Pro Asn Ser	
425 430 435	
ACC ATG GTA GGC ACG GGC ATG GCC CCC GAG GGC ATC AGC CAG AAC GAA	1457
Thr Met Val Gly Thr Gly Met Ala Pro Glu Gly Ile Ser Gln Asn Glu	
440 445 450	
GTG GTC TAT TCC CTC ATG GCT GAG CTG GGC TGG CGA AAG GAC CCA GTG	1505
Val Val Tyr Ser Leu Met Ala Glu Leu Gly Trp Arg Lys Asp Pro Val	
455 460 465	
CCA GAT TTG GCA GCC TGG GTG ACC AGC TTT GCC GCC CGG CGG TAT GGG	1553
Pro Asp Leu Ala Ala Trp Val Thr Ser Phe Ala Ala Arg Arg Tyr Gly	
470 475 480	
GTC TCC CAC CCG GAC GCA GGG GCA GCG TGG AGG CTA CTG CTC CGG AGT	1601
Val Ser His Pro Asp Ala Gly Ala Ala Trp Arg Leu Leu Leu Arg Ser	
485 490 495 500	

GTG TAC AAC TGC TCC GGG GAG GCC TGC AGG GGC CAC AAT CGT AGC CCG	1649		
Val Tyr Asn Cys Ser Gly Glu Ala Cys Arg Gly His Asn Arg Ser Pro			
505	510	515	
CTG GTC AGG CGG CCG TCC CTA CAG ATG AAT ACC AGC ATC TGG TAC AAC	1697		
Leu Val Arg Arg Pro Ser Leu Gln Met Asn Thr Ser Ile Trp Tyr Asn			
520	525	530	
CGA TCT GAT GTG TTT GAG GCC TGG CGG CTG CTG CTC ACA TCT GCT CCC	1745		
Arg Ser Asp Val Phe Glu Ala Trp Arg Leu Leu Leu Thr Ser Ala Pro			
535	540	545	
TCC CTG GCC ACC AGC CCC GCC TTC CGC TAC GAC CTG CTG GAC CTC ACT	1793		
Ser Leu Ala Thr Ser Pro Ala Phe Arg Tyr Asp Leu Leu Asp Leu Thr			
550	555	560	
CGG CAG GCA GTG CAG GAG CTG GTC AGC TTG TAC TAT GAG GAG GCA AGA	1841		
Arg Gln Ala Val Gln Glu Leu Val Ser Leu Tyr Tyr Glu Glu Ala Arg			
565	570	575	580
AGC GCC TAC CTG AGC AAG GAG CTG GCC TCC CTG TTG AGG GCT GGA GGC	1889		
Ser Ala Tyr Leu Ser Lys Glu Leu Ala Ser Leu Leu Arg Ala Gly Gly			
585	590	595	
GTC CTG GCC TAT GAG CTG CTG CCG GCA CTG GAC GAG GTG CTG GCT AGT	1937		
Val Leu Ala Tyr Glu Leu Leu Pro Ala Leu Asp Glu Val Leu Ala Ser			
600	605	610	
GAC AGC CGC TTC TTG CTG GGC AGC TGG CTA GAG CAG GCC CGA GCA GCG	1985		
Asp Ser Arg Phe Leu Leu Gly Ser Trp Leu Glu Gln Ala Arg Ala Ala			
615	620	625	
GCA GTC AGT GAG GCC GAG GCC GAT TTC TAC GAG CAG AAC AGC CGC TAC	2033		
Ala Val Ser Glu Ala Glu Ala Asp Phe Tyr Glu Gln Asn Ser Arg Tyr			
630	635	640	
CAG CTG ACC TTG TGG GGG CCA GAA GGC AAC ATC CTG GAC TAT GCC AAC	2081		
Gln Leu Thr Leu Trp Gly Pro Glu Gly Asn Ile Leu Asp Tyr Ala Asn			
645	650	655	660
AAG CAG CTG GCG GGG TTG GTG GCC AAC TAC TAC ACC CCT CGC TGG CGG	2129		
Lys Gln Leu Ala Gly Leu Val Ala Asn Tyr Tyr Thr Pro Arg Trp Arg			
665	670	675	
CTT TTC CTG GAG GCG CTG GTT GAC AGT GTG GCC CAG GGC ATC CCT TTC	2177		
Leu Phe Leu Glu Ala Leu Val Asp Ser Val Ala Gln Gly Ile Pro Phe			
680	685	690	
CAA CAG CAC CAG TTT GAC AAA AAT GTC TTC CAA CTG GAG CAG GCC TTC	2225		
Gln Gln His Gln Phe Asp Lys Asn Val Phe Gln Leu Glu Gln Ala Phe			
695	700	705	
GTT CTC AGC AAG CAG AGG TAC CCC AGC CAG CCG CGA GGA GAC ACT GTG	2273		
Val Leu Ser Lys Gln Arg Tyr Pro Ser Gln Pro Arg Gly Asp Thr Val			
710	715	720	

GAC CTG GCC AAG AAG ATC TTC CTC AAA TAT TAC CCC GGC TGG GTG GCC	2321
Asp Leu Ala Lys Lys Ile Phe Leu Lys Tyr Tyr Pro Gly Trp Val Ala	
725 730 735 740	

GGC TCT TGG TGATAGATTGCCACCACTGGCCTTGTTTCCCGCTAATT	2370
Gly Ser Trp	

CCAGGGCAGA TTCCAGGGCC CAGAGCTGGA CAGACATCAC AGGATAACCC AGGCCTGGGA 2430

GGAGGCCCA CGGCCTGCTG GTGGGGCTG ACCTGGGGGG ATTGGAGGGA AATGACCTGC 2490

CCTCCACCAC CACCCAAAGT GTGGGATTAA AGTACTGTTT TCTTTCCACT TAAAAAAA 2550

AAAAAAAGTCG AGCGGCCGCG AATTC 2575

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 743 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(ix) FEATURE:

- (A) NAME/KEY: Potentially-glycosylated Asn site,
- (B) LOCATION: 261

(ix) FEATURE:

- (A) NAME/KEY: Potentially-glycosylated Asn site,
- (B) LOCATION: 272

(ix) FEATURE:

- (A) NAME/KEY: Potentially-glycosylated Asn site,
- (B) LOCATION: 435

(ix) FEATURE:

- (A) NAME/KEY: Potentially-glycosylated Asn site,
- (B) LOCATION: 503

(ix) FEATURE:

- (A) NAME/KEY: Potentially-glycosylated Asn site,
- (B) LOCATION: 513

(ix) FEATURE:

- (A) NAME/KEY: Potentially-glycosylated Asn site,
- (B) LOCATION: 526

(ix) FEATURE:

- (A) NAME/KEY: Potentially-glycosylated Asn site,
- (B) LOCATION: 532

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Glu Ala Val Ala Val Ala Ala Val Gly Val Leu Leu Leu Ala  
1 5 10 15

Gly Ala Gly Gly Ala Ala Gly Asp Glu Ala Arg Glu Ala Ala Ala Val  
20 25 30

Arg Ala Leu Val Ala Arg Leu Leu Gly Pro Gly Pro Ala Ala Asp Phe  
35 40 45

Ser Val Ser Val Glu Arg Ala Leu Ala Ala Lys Pro Gly Leu Asp Thr  
50 55 60

Tyr Ser Leu Gly Gly Gly Ala Ala Arg Val Arg Val Arg Gly Ser  
65 70 75 80

Thr Gly Val Ala Ala Ala Ala Gly Leu His Arg Tyr Leu Arg Asp Phe  
85 90 95

Cys Gly Cys His Val Ala Trp Ser Gly Ser Gln Leu Arg Leu Pro Arg  
100 105 110

Pro Leu Pro Ala Val Pro Gly Glu Leu Thr Glu Ala Thr Pro Asn Arg  
115 120 125

Tyr Arg Tyr Tyr Gln Asn Val Cys Thr Gln Ser Tyr Ser Phe Val Trp  
130 135 140

Trp Asp Trp Ala Arg Trp Glu Arg Glu Ile Asp Trp Met Ala Leu Asn  
145 150 155 160

Gly Ile Asn Leu Ala Leu Ala Trp Ser Gly Gln Glu Ala Ile Trp Gln  
165 170 175

Arg Val Tyr Leu Ala Leu Gly Leu Thr Gln Ala Glu Ile Asn Glu Phe  
180 185 190

Phe Thr Gly Pro Ala Phe Leu Ala Trp Gly Arg Met Gly Asn Leu His  
195 200 205

Thr Trp Asp Gly Pro Leu Pro Pro Ser Trp His Ile Lys Gln Leu Tyr  
210 215 220

Leu Gln His Arg Val Leu Asp Gln Met Arg Ser Phe Gly Met Thr Pro  
225 230 235 240

Val Leu Pro Ala Phe Ala Gly His Val Pro Glu Ala Val Thr Arg Val  
245 250 255

Phe Pro Gln Val Asn Val Thr Lys Met Gly Ser Trp Gly His Phe Asn  
260 265 270

Cys Ser Tyr Ser Cys Ser Phe Leu Leu Ala Pro Glu Asp Pro Ile Phe  
275 280 285

Pro Ile Ile Gly Ser Leu Phe Leu Arg Glu Leu Ile Lys Glu Phe Gly  
290 295 300

Thr Asp His Ile Tyr Gly Ala Asp Thr Phe Asn Glu Met Gln Pro Pro  
305 310 315 320

Ser Ser Glu Pro Ser Tyr Leu Ala Ala Ala Thr Thr Ala Val Tyr Glu  
325 330 335

Ala Met Thr Ala Val Asp Thr Glu Ala Val Trp Leu Leu Gln Gly Trp  
340 345 350

Leu Phe Gln His Gln Pro Gln Phe Trp Gly Pro Ala Gln Ile Arg Ala  
355 360 365

Val Leu Gly Ala Val Pro Arg Gly Arg Leu Leu Val Leu Asp Leu Phe  
370 375 380

Ala Glu Ser Gln Pro Val Tyr Thr Arg Thr Ala Ser Phe Gln Gly Gln  
385 390 395 400

Pro Phe Ile Trp Cys Met Leu His Asn Phe Gly Gly Asn His Gly Leu  
405 410 415

Phe Gly Ala Leu Glu Ala Val Asn Gly Gly Pro Glu Ala Ala Arg Leu  
420 425 430

Phe Pro Asn Ser Thr Met Val Gly Thr Gly Met Ala Pro Glu Gly Ile  
435 440 445

Ser Gln Asn Glu Val Val Tyr Ser Leu Met Ala Glu Leu Gly Trp Arg  
450 455 460

Lys Asp Pro Val Pro Asp Leu Ala Ala Trp Val Thr Ser Phe Ala Ala  
465 470 475 480

Arg Arg Tyr Gly Val Ser His Pro Asp Ala Gly Ala Ala Trp Arg Leu  
485 490 495

Leu Leu Arg Ser Val Tyr Asn Cys Ser Gly Glu Ala Cys Arg Gly His  
500 505 510

Asn Arg Ser Pro Leu Val Arg Arg Pro Ser Leu Gln Met Asn Thr Ser  
515 520 525

Ile Trp Tyr Asn Arg Ser Asp Val Phe Glu Ala Trp Arg Leu Leu Leu  
530 535 540

Thr Ser Ala Pro Ser Leu Ala Thr Ser Pro Ala Phe Arg Tyr Asp Leu  
545 550 555 560

Leu Asp Leu Thr Arg Gln Ala Val Gln Glu Leu Val Ser Leu Tyr Tyr  
565 570 575

Glu Glu Ala Arg Ser Ala Tyr Leu Ser Lys Glu Leu Ala Ser Leu Leu  
580 585 590

Arg Ala Gly Gly Val Leu Ala Tyr Glu Leu Leu Pro Ala Leu Asp Glu  
595 600 605

Val Leu Ala Ser Asp Ser Arg Phe Leu Leu Gly Ser Trp Leu Glu Gln  
610 615 620

Ala Arg Ala Ala Ala Val Ser Glu Ala Glu Ala Asp Phe Tyr Glu Gln  
625 630 635 640

Asn Ser Arg Tyr Gln Leu Thr Leu Trp Gly Pro Glu Gly Asn Ile Leu  
645 650 655

Asp Tyr Ala Asn Lys Gln Leu Ala Gly Leu Val Ala Asn Tyr Tyr Thr  
660 665 670

Pro Arg Trp Arg Leu Phe Leu Glu Ala Leu Val Asp Ser Val Ala Gln  
675 680 685

Gly Ile Pro Phe Gln Gln His Gln Phe Asp Lys Asn Val Phe Gln Leu  
690 695 700

Glu Gln Ala Phe Val Leu Ser Lys Gln Arg Tyr Pro Ser Gln Pro Arg  
705 710 715 720

Gly Asp Thr Val Asp Leu Ala Lys Lys Ile Phe Leu Lys Tyr Tyr Pro  
725 730 735

Gly Trp Val Ala Gly Ser Trp  
740

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10380 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo sapiens

(viii) POSITION IN GENOME:

- (A) CHROMOSOME SEGMENT: Chromosome 17

(ix) FEATURE:

- (A) NAME/KEY: exon 1
- (B) LOCATION: 990..1372

(ix) FEATURE:

- (A) NAME/KEY: exon 2
- (B) LOCATION: 2115..2262

(ix) FEATURE:  
(A) NAME/KEY: exon 3  
(B) LOCATION: 3056..3202

(ix) FEATURE:  
(A) NAME/KEY: exon 4  
(B) LOCATION: 3387..3472

(ix) FEATURE:  
(A) NAME/KEY: exon 5  
(B) LOCATION: 5667..5923

(ix) FEATURE:  
(A) NAME/KEY: exon 6  
(B) LOCATION: 7745..8955

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ATAATGAGCA GTGAGGACGA TCAGAGGTCA CCTTCCTGTC TTGGTTTG GAGGTTTGA 60  
CCAGTTTCTT TGCTGCATTC TGTTTTATCA GCGGGGTCTT GTGACCTTTT ATCTTGTGCT 120  
GACCTCTGT CTCATCCTGT GACGAAGGCC TAACCTCTG GGAATTCA GCAGCAGGTC 180  
TCTGCCTCAT TTTACCCAGC CCCTGTTCAA GATGGAGTCG CTCTGGTTGG AAACCTCTGA 240  
CAAATGACA GCTCCTGTTA TGTTGCTGCT GCTGCCGCCA ATGGACAGCC TTTAACGTGC 300  
CCGCCAGCCC TGCTCCACCG CGGGCCTGGG CTCACATGGC CCCATCCCTC CTCGAACCTC 360  
CTAGCCTGTT AGTTACTCAA ATCTGCAAGC TCTCTGCCTT CTCAGGGCCT TCAATAATG 420  
CATTTCTTCT GTCTGGAAGG CTCTTCTTT CCCTCTCTA GCCAATTCCCT ATTCAATCCCT 480  
GAGTTTCAGA TAAAAAGTC CTTCTTTGG AAACCTTACT TCGCTACTTC GCTACTTACT 540  
GCACTACTTC GCAGCATCAC AACTATGATG GAAATCCTTA CTTACGTTAA ATATCTGGTT 600  
TCTAGGTAC CTCCTGACG GGGACGGTAG GGACCGTCTT CTCGTTCATC AGTAGGGAAG 660  
TAGCTATGGC AGTGCTGTAC AAAAAATAA CTCCAAATGT GTATTTATTA GATGGTTGGA 720  
TGGAAAGTTAT TTGCGTGTGA AAGCGCGTTT TACCCGAAGG CGCTCTGTGA GGGCCAGCGG 780  
GTCCTCCCTCG GCCCTGGAGC CGGGGTCAACA CGCTCCCCAC CGCGTGCCTG CACGAGACGC 840  
CCCCAAGGGA GTATCCTGGT ACCCGGAAGC CGCGACTCCT GGCCCTGAGC CGGGGCTTAG 900  
CCTTCGGGTC CACGTGGCCG GAGCCGGCAG CTGATTGGAC GCGGGCCGCC CCACCCCTG 960  
GCCGTGCGGG GACCCGCAGG ACTGAGACCA TGGAGGCGGT GGCGGTGGCC GCGGCGGTGG 1020  
GGGTCCCTCT CCTGGCCGGG GCGGGGGCGC CGGCAGGCGA CGAGGCCCGG GAGGGCGGG 1080  
CCGTGCGGGC GCTCGTGGCC CGGCTGCTGG GGCCAGGCC CGCGGGCGAC TTCTCCGTGT 1140

CGGTGGAGCG CGCTCTGGCT GCAAGCCGG GCTTGGACAC CTACAGCCTG GCGGGCGCG 1200  
GCGCGGCGCG CGTGCCTGGTG CGCGGCTCCA CGGGCGTGGC GGCGCGCGCG GGGCTGCACC 1260  
GCTACCTCTCG CGACTCTCTGT GGCTGCCAG TGCCCTGGTC CGGCTCTCAG CTGCGCCCTG 1320  
CGCGGCCACT GCCAGCCGTG CGGGGGGAGC TGACCGAGC CACGCCAAC AGGTACCGCC 1380  
CCGAAGCTTC CCCGCGTCCG CGCGAGGCGC TTACCCCCCTC CGCGAGGCCG TGCCACCCAA 1440  
ATCGGGGAGGC TGAGCGGGGA GCGCTGGCG GAAGGCCAG CTGCGCCGCC TCCAGCAGCT 1500  
GTGTGGCCTT GAGCCAGCCA CTCTGCCTTT CAGAGCCTCG GCTGGCCAC CTGAAAAACG 1560  
GAAAGAAGAC GCCTACCGTG CAGTGTATT GTGAGGATTT GCACGATGAT GGGCATAGAA 1620  
TTTGTGGTGC ACAATTGGTG ATGAGTGAAT TTTCTTGCT TCCTCCCCCA CCTTCTCTTT 1680  
GAACCTGCGG ACTGAGGAAG GACGCCCTCA TCCCCCACCC TACAGGCCG TGTTCCAGCG 1740  
CCTGCCACAC TATGGAGTGA TGTGTTACCA CAGCTGTCTT CCCCTGCCA TCTGTTAGAC 1800  
TGTGGGGGCA GGGATTCCCC GTTCCAGGAA AACACCGTGC AGAGGAGGGG CTCTGGCAGT 1860  
GTGGCATGAA AGTGGAAAT GCCACCCAAA TACCCGCCAG GCTAGAGGGC CCTGGGAGAG 1920  
TGCAGGGGAC GAGTGCCTCA GAAGCCCAGC CCCGGTACCT GGTCTCAGCT CCACCTGGGG 1980  
TGGGTCCCAG TGTGCAGCAG AAGGGCCGAG TTTGGAGCCC CTCCCCCTC CTCTAGGTGG 2040  
GGGATGGGGG ATTTGTTCCA GGGCGTGGA CCCTCCAGGG TGGGATGCGC CCCTGCTCAT 2100  
GACACTGCC GCAGGTACCG CTATTACCA AATGTGTGCA CGCAAAGCTA CTCTTCTGT 2160  
TGGTGGGACT GGGCCCGCTG GGAGCGAGAG ATAGACTGGA TGGCGCTGAA TGGCATCAAC 2220  
CTGGCACTGG CCTGGAGCGG CCAGGAGGCC ATCTGGCAGC GGGTGCCTGC CCACTGTCCC 2280  
TTCCCCACCC TCCTCTATGG CGGGAGCCAC CGTAGGTGTT TTCACCCGCC CCCCAGCATG 2340  
GGCGCAGTGT CTCTCTCTAG AAGTGTCTTC AGCGTGCACA GTGGCTTGGG CCTCTCTAAA 2400  
ACTGAGGCTT CCGGCCGGGC GCGGTGGCTC ACGCCTGTCA TCCCAGCACT TCGGGAGGCC 2460  
TAGGCAGGCG GATCAGGAGT TCAGGAGATC GAGACCATCC TGGCCAACAT TGTGAAACCC 2520  
CGTCTCTACT AAAATACAAA GAAATAGCAA CCTGGGCAAC AGAGCGAGAC TCTGTCTAAA 2580  
AAAAAAAAAAA AAAAAAAACTG AGGCTTCCAG TTTGAGGAGT GGGGCTCTT CCCCCATCTC 2640  
CCCTATGCAG CCAATCACCT GGTCCTTGG ATCCAACCTA TGGGCAGCTC TAGATCTGCC 2700  
TCCCTGGAAAG CTTCTGTGCT GCAATGGCTG CTCCAGGCTC TGCTTAAGCT CTTCACACAG 2760  
TTGCCCTGCC CTTCCATCTG GCACTCTTGC TCCATGAAGC CTTCTAAGGC CTTCTGTG 2820

GGGGAAAGCC CCTTTGTGCC CCATCTCTC ACCCATGCAGA CAAAGGCAAC ACAGTGAAC 2880  
CACCTACTCA CAGGTCTCTT TCCCTCTGGC TGTGGGCTCC TTGATGGCAG CGTCGGATT 2940  
TTGTCTCAGT AGCCTAGCA CCCAGCACAA AGAAGCAATG AGTGAATGGT TGTGAAATGA 3000  
ATGAATGAAT GAATGAAGAT GAATATATTT CTATGTGTGG GCCCTTCTTC CTCAGGTGTA 3060  
CCTGGCCTTG GGCCTGACCC AGGCAGAGAT CAATGAGTTC TTTACTGGTC CTGCCTTCCT 3120  
GGCCTGGGG CGAATGGGCA ACCTGCACAC CTGGATGGC CCCCTGCCCT CCTCCTGGCA 3180  
CATCAAGCAG CTTTACCTGC AGGTAAAAGG ATGGAAAAGG GAAGGGCAG AATCGGTGAT 3240  
AGATGGTCAT GGGCCCAGGA AGGGTGGTAT TAGGCCGGCC CCAGGGCTCT TAACTGAGGC 3300  
GGGGGCTGC GTGTATCCTG GGAGATGAGG GCCTTCTCAT AGGACAGCAG TGGCCATGCT 3360  
CACCACCCCTT CCTTCTGTT CTCAGCACC GGGCCTGGAGA CCAGATGCGC TCCTTCGGCA 3420  
TGACCCAGT GCTGCCTGCA TTGCGGGGC ATGTTCCCGA GGCTGTCACC AGGTGAGGTT 3480  
CCGCTCACCC CCTCCACTTA GCTCAGAGAG GGAATTTAT TCCCTTCTAG AACATGACTT 3540  
AAAAACTTAA GCTCTGGGCC GGGCGCAGTG GCTCACGCCT GTAATCCCAG CACTTTGGGA 3600  
GGCGAGTTG GCGGGATCAC CTGAGGTCAG GAGTTCGAGA CCAGCCTGGC CAACATGGTG 3660  
AAACCTGTC TCTACTAAAA ATATAAAAAT TAGCTGGCA TGTTGGCAGC CGCCCTGTAAT 3720  
CCCATCTACT TAGGAGGCTG AGACAGGAGA ATTGCTTAAA CCTGGGAGGC AGACGTTGCA 3780  
GTGAGTCAG ATCACGCCAT TGCACCTCAG CCTGGGTGAC GAGCGAAACT CTGTCTCAA 3840  
CAAACAAACA AGCTCTGGAC GTAGGCCTGG GTTTGATTTC TGACTCTGCT ACTAATTAGC 3900  
TGTGTGACTT CGGGCAGATG ACATGACTGC TCTGTGCCTC AGTTTCCCTTA CTGTAAAAT 3960  
GGGATCTCTA CCCACTTCGC TGTAGGGTTT GTAATTATCT CTCGATCTAT CTGTGACTTT 4020  
GCACAGAGTG CTAGCAAATG GCAGCCTCTG GGAGTGGCAG CAGGGGTGCT CCAGTGTCCC 4080  
TTGTCCCTCC TGTTCCCTCTG TGCTTCCAG CCATCCTCTC ACATGTGGTT GGGAAAAGTC 4140  
TTCAAGGCTC ACCTGAGACC TCCCTCTCTT CAGGAAGCCT TGCTAGTGCC CCGCATGACC 4200  
TCCCTTGAC CTGCTAATGT CTGGCTCCCA TACTCTCGTA GGACTTAATG CATGCCAGTG 4260  
GCCTCCCTGC CGCCCTCTTT GCCCCCATCA CCAGGTGGCA GGAAACTCAC TCATTCATTC 4320  
AATAAAACTTG GTCCAGCTGT CTGAGGCTGC CAGAACTGGC TGTGCTGGGT CCTGGGAGGC 4380  
GGCAAGAAAG GTGCCAAGG GCTTACCCCT GATAGGAGAG ATATGTTGGC TGAAGGATAC 4440  
AATGTGGGGA CAAGGACAGG AATATATGTG GGTTCCGCTC TCCTCTGCCG GGAGAGAGGG 4500

GCAGGAAGGG CTCAGGGCAG AGCCCAGCCT TGAAAAATGA GTGTTGCTTG GACGGACGCT 4560  
TGGCTAATGC TTGTAATCCT AGCGTTTTGG GAGGCTGAGG CGTATGGATC ACCTGCGGTC 4620  
AGGAGTTAAA GACCAGCCCTG GCCAACATGG CGAACACCCCA TCTCTACTAA AAGTACAAAA 4680  
ATTAGCCAGG CGTGGTGGCG GGCTCCTGTA ATCCCAGCTA CTCGGTAGGC TGAGGCATGA 4740  
GAATCTCTTG AAGCCAGGGG CCAGAGACTG CAGTGAGCCG AGATCACACC ACITTCACTCC 4800  
AGCCTGGGTG ACAGAGTGTAG ACTCCGTCTC AAAAAAAA AAAAAAAAG GAAAGAAAAT 4860  
TAAACACCTC ATGTTCTCAC TCATAGTGGG AGTTGAACAA TGAGAACAAAC ATGGACACAG 4920  
GAAGGGGAAC ATCACACACC GGGGCCTTTC GCGGTGTGGG GGTCAAGGGG AGGAGTAGCA 4980  
TTGGGACAGA TACTTAATGC ATGCGGGGCT GAAAACCTAG ATGATGGGTT GATGGGTGCA 5040  
GCAAACCACCA ATGGCACATG TATACTATG CAACAAACCT GCATGTTCTG CACAGAACTG 5100  
AACTGAAAGT ATAATTAAAA AAAAAAAA AAGCTGGGTG CGGTGGCCCA CACCTGTAAT 5160  
CCCAGCACTT TGGGAGGGCG AGACGGGCGG ATCACAAGGT CACGAGATCG AGACCATCCT 5220  
GGCTAACACA GTGAAACTCA GTCTCTACTA AAAATACAAA AAATTAGCCG GGTGTGGTGG 5280  
CGGGCACCTG TAGTCCCAGC TACTAGGGAG GCTGAGGCAG GAGAATGGCA TGAACCTGGG 5340  
AGGCAGAGCT TGCACTGAGC TGAGAACATGC CCACTGCACT CCAGCCTGGG GGACAGAGTG 5400  
AGACTCTGCC TCAAAAAAAA AAAAAAAAG AAAGAAAAG GAGCGTTGCT TGTTTCAGGC 5460  
CACAGGAAGG GGAGAGATAG TGAAAGTTT TCAGAGAAGG TGGCCAGGGA AGGAGAAGAA 5520  
AGGACTGTAG GCAGAGAGCA TAGCCTGTAC AAAGCCATAG AGGCAAGAGA AACCAAGGAGC 5580  
TGTAGAGAAG TTGGCAAGGC TGTTGAACAC TATGGTGAAC ACTATGGCGG CTTCCATGAA 5640  
ATATCTGAGC TTTGCTCCC CACTAGGGTG TTCCCTCAGG TCAATGTCAC GAAGATGGGC 5700  
AGTTGGGCC ACTTTAACTG TTCCCTACTCC TGCTCCTTCC TTCTGGCTCC GGAAGACCCC 5760  
ATATCCCCA TCATCGGGAG CCTCTTCTCG CGAGAGCTGA TCAAAGAGTT TGGCACAGAC 5820  
CACATCTATG GGGCCGACAC TTTCAATGAG ATGCAGCCAC CTTCCCTCAGA GCCCTCCTAC 5880  
CTTGCGCGAG CCACCACTGC CGTCTATGAG GCCATGACTG CAGGTACAGT GCCTGGGTGG 5940  
GGTGGGAGAG CCCCCCAGAC CCTCAAAAG AAGGGAGTAG CAGATGTCAG TAGGGTAGG 6000  
CAGAGGGACT GGAATAATGC CTCGCCATAA CACACAGTAC TTTATAGTTT ACCAACGACCG 6060  
TGTACACATG CGTTGTCTCA GTGAATCCCA CTGTTGGTTGA GAGGTGAGCT CTGGAAGCCA 6120  
ACAACCTGGG TCACACCTCG CGCTCCTATT TCCTGGCCGT GTGACTTATG ACTCATGACC 6180

TCCTTCCAG TGTCTCGTT GCTTTCTG TAAACTGGGA CTACCTCATA GGTAGAATAA 6240  
CGCCTGGCCC AGAGCAAAGG CCACTAAGAG CTAGCTATGA ACAAGGATTG TGTTTCATCT 6300  
CTGCGTGGTT GCTGAAGTAG GCACTCAGG CAGGAGGTGA GTGGATGTGC CTAAAGGCAC 6360  
TAAGTGCAGCA TCCTGCTACA AAACGTGAA GCCAGGGCTC CTTCTGCCA CTTAAAGGAG 6420  
GAGTGGAGCA GAGGGCGCCC AAGTCAGGAA TGACTTAGTG GAGAGGCGTC TGTGTTGGCC 6480  
AGGAAGGAA CAGATCAGCT CAGCCTTCT TGAGCAGTAC TGCTCCAAGT GTGACCCCCA 6540  
ACCAAGCAGCA GCAGCAGCAG CAGCCGAGC TGTGAGATGG CAAATTCTCA GGCCCTACCC 6600  
AAGACACTGAA GGAGAAGCTA CATTTTTTTT TTTTTTGAGA CAGATTTCAC TCTGTTGCTG 6660  
AGGCTGGAGC ACAGTGGCAC AATCTCATCT CACTGCAACC TTGCTCTCCT AGGTTCAAGC 6720  
GATTCTCCTG CCTCAGCCCTC CCGAGTAGCT GGGACTATAG GCACCCGCCA CCACGCCGG 6780  
CAATTTTGTG TTGTTTTGAG ATAGAGTCTC GCTCTGTAC CCAGGCTGGA GTGCAGTGGC 6840  
ACGATCTCAG TTCACTGCAA CCTCTGCTTC CTGAGTTCAA GCGATTCTCC TGCCCTCAGCC 6900  
TCCTGAGTAG CTGGGATTAC AGGCAGCCCC CAACCACACT CGGCTAATTG TTGTATTTT 6960  
AGTAGAGACG GGGTTTCGCT ATGTTAGGTCA AGCTGGTTTC AAACCTCTGA CCTCAAATGA 7020  
TTCGCCCACT TCAGCCTCCC AAAGTGTGG GATTACAGGT GTGAGGCCACC TTGCTGGCC 7080  
AATTTTGTA TTTTTAGTAG AAACAGGTTT CACCATGGTG GCCAGACTGG TCTCAAATC 7140  
CTGACCTCAG GTGAACCTGCC CACCTCAGCC TCCCAAAGTA CTGGTATTAC AGGCGTGTAC 7200  
CACTGCACT GGCCTTGATT TTGTTTTGAA GACAGAACTCT TACTCTGTGCCCAGACTGG 7260  
AGTGCACTGG CACAATCTCA GCTCACTGCA ACTTCTGCCT CATGGGTCA AGTGATTCTT 7320  
GTGCCCTCTAC CTCCCGAGTA GCCGGGATTAA CAGGCACCTG CCATTACGCT AGGCTAATTG 7380  
TTGTATTTT AGTATAGACA GGGTTTCCCC ACATTGGCCA GGCTGGTCTG GAACTCCTGG 7440  
GCTCAAGTGA TCCACCTGCT TCAGCCCTC AGAGTACTGG GATTATAGGT GTGGGCCACC 7500  
ACGCCCATTC AGAAACCTCC ATGTTTTAAG GAGCCCTCTG GGTAACCTCTC ATGTTCAACCC 7560  
AAGCTGCTGA ACCCTGTCTT GGAGTTTCA GAGGGACGCG TATGTGCCAC AGAGCGTCCC 7620  
GCTGGTGGGG GTCATGGGAA GCCATGACCT GGGATAGACA GTCTGTGTA GAGTGGGGTG 7680  
AACATCCCCCTT GGGCCCTCTG TTTCATCACT CCTCTCTCTC GTTCCCCCTA CCTCTGTCC 7740  
ACAGTGGATA CTGAGGCTGT GTGGCTGCTC CAAGGCTGGC TCTTCCAGCA CCAGCCGAG 7800  
TTCTGGGGGC CGGCCAGAT CAGGGCTGTG CTGGGAGCTG TGCCCCGTGG CGCCCTCCTG 7860

GTTCTGGACC TGTGTTGCTGA GAGCCAGCCT GTGTATAACCC GCACTGCCTC CTTCCAGGGC 7920  
CAGCCCTTCA TCTGGTGCAT GCTGCACAAC TTTGGGGAA ACCATGGTCT TTTTGGAGCC 7980  
CTAGAGGCTG TGAACGGAGG CCCAGAACGCT GCCCGCCTCT TCCCCAACTC CACCATGGTA 8040  
GGCACGGGCA TGGCCCCGA GGGCATCAGC CAGAACGAAG TGGTCTATTG CCTCATGGCT 8100  
GAGCTGGGCT GGCAGAAAGGA CCCAGTGCCA GATTTGGCAG CCTGGGTGAC CAGCTTGCC 8160  
GCCCGGCGGT ATGGGGTCTC CCACCCGGAC GCAGGGGAG CGTGGAGGCT ACTGCTCCGG 8220  
AGTGTGTACA ACTGCTCCGG GGAGGGCCTGC AGGGGCCACA ATCGTAGCCC GCTGGTCAGG 8280  
CGGGCGTCCC TACAGATGAA TACCAAGCATC TGGTACAACC GATCTGATGT GTTTGAGGCC 8340  
TGGCGGCTGC TGCTCACATC TGCTCCCTCC CTGGCACCA GCCCCGCCTT CCGCTACGAC 8400  
CTGCTGGACC TCACTCGGCA GGCAGTGCAG GAGCTGGTCA GCTTGTACTA TGAGGAGGCA 8460  
AGAAGCGCCCT ACCTGAGCAA GGAGCTGGCC TCCCTGTTGA GGGCTGGAGG CGTCCTGGCC 8520  
TATGAGCTGC TGCCGGCACT GGACGGAGTG CTGGCTAGTG ACAGCCGCTT CTTGCTGGC 8580  
AGCTGGCTAG AGCAGGCCG AGCAGCGGCA GTCAGTGAGG CCGAGGCCGA TTTCTACGAG 8640  
CAAACAGGCC GCTACCAGCT GACCTTGTGG GGGCCAGAAC GCAACATCCT GGACTATGCC 8700  
AACAAAGCAGC TGGCGGGGTT GGTGGCCAAC TACTACACCC CTCGCTGGCG GCTTTCCCTG 8760  
GAGGCCTGG TTGACAGTGT GGCCAGGGC ATCCCTTCC AACAGCACCA GTTTGACAAA 8820  
AATGTCCTCC AACTGGAGCA GGCCTTCGTT CTCAGCAAGC AGAGGTACCC CAGCCAGCCG 8880  
CGAGGAGACA CTGTGGACCT GGCCAGAACAG ATCTTCCCTCA AATATTACCC CGGCTGGGTG 8940  
GCCGGCTCTT GGTGATAGAT TCGCCACAC TGGGCCTTGT TTTCCGCTAA TTCCAGGGCA 9000  
GATTCCAGGG CCCAGAGCTG GACAGACATC ACAGGATAAC CCAGGGCTGG GAGGAGGCC 9060  
CACGGCCTGC TGGTGGGGTC TGACCTGGGG GGATTGGAGG GAAATGACCT GCCCTCCACC 9120  
ACCACCCAAA GTGTGGGATT AAAGTACTGT TTTCTTCCA CTTAAACTGA TGAGTCCCT 9180  
GGGTCTGTCA AAATGAGAACAG GTCACTGCTG CCACGCTTGG GAGGAGTCAG GGCTATAGCA 9240  
TGGCCCTGGG GTGGGACCTG TTCTCCCATC CCTTGCCCTCA CGTCCCTGTT TTTGTTGTT 9300  
TGTTTGTGTTG TGACGGAGCC TTGGTCTGTT GCCCAGGCTT GAGTACAATG GCACAGTCTC 9360  
GGCTCACTGC AACCTCCGCC TCCTGGGTTA AAGCAATTCT TGTGCCTCAG CCTCCCCGGT 9420  
AGCTGGGACT ATAGGCATGC ACCACCACAC CAGGCTAATT TTTTTTTTC CAAGATGGAG 9480  
TCCTGCTCTG TCGCCCAGGT TGGAGTTAG TGGCACCATA TTGGTTTACT GCAACCTCTG 9540

CCTCCGGGT TCAAGCAATT CTCCCTGCCTC AGTCTACAG GGAGTTAGGA CTACGGGCCT 9600  
GTGCCATCAC GCCTGGCTAA TTTTGTTATT TTTCATAGAG ATAAGGTTTC ACCATGTTGG 9660  
CCAGGCTGGT CTTTAACCTCC TGAACCTCAAG TGATCCACCT GCCTCGGCCT TCCAAAGTGC 9720  
TGGGATTACA GGAGTGAGCC ACCGTGCCCG GCCATGTCTC TCTTTTTAAC ACTAATGTTA 9780  
CCCTGACCTT TGAACGTAGA ATGCCCTTCT GTTGCAGGAA AACCTCTTTT CAAACCATGT 9840  
TTGTCCTTTG CTGGCATGCC ACAGAACAG TCACCAACAC AGAAGACTTC TGTGACCAAA 9900  
TATTTGGAGG ATTTTCCCCA CACACACAA GCAGCAGACA TCAGCTGGGT GTCCTCCAAT 9960  
TCAGTTCCA TGTAATCAAC CAGAGACAGC ATCAGATCCC ACAGGGTTAG GGTGCAGATC 10020  
CATGAGACCA CCCCCTCCTT CCCAACGGTT ACAAGTCCTG ATCCCTGGAA CTTCTGACTA 10080  
ACTGGCTTCA AGTTGGAGTT CCCATGACCC CCTTCCCCCTC TTTGGAGTCA ACTCATTTGC 10140  
GACAGTGACC CACGAAACAC AGGGAAACCC TTATTATGTT TATTGCTTTA TTACAGAGGA 10200  
AAAAAAATTTT TTTCCTTCTT TTTTGAGACA GGGTCTCACT CTGTCATCCA GAATGACTGC 10260  
AGTGGCAGGA TCTGGCTCCG TCACCCAGGC TGGAGTGCAG TGGCATGATC TCGGCTCACT 10320  
ACAGCCTCCA TCCCCCCCCA ACCCCACGCC TCAGCGCCCC ACCCCGCAAG TGGCTGGGAC 10380

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (iii) HYPOTHETICAL: NO
- (v) FRAGMENT TYPE: N-terminal

(vi) ORIGINAL SOURCE:  
(A) ORGANISM: Homo sapiens

(ix) FEATURE:  
(A) NAME/KEY: Modified-site  
(B) LOCATION: 10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Asp Glu Ala Arg Glu Ala Ala Ala Val Arg Ala Leu Val Ala Arg  
1 5 10 15

Leu Leu Gly Pro Gly  
20

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 18 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: N-terminal

(vi) ORIGINAL SOURCE:  
(A) ORGANISM: Homo sapiens

(ix) FEATURE:  
(A) NAME/KEY: Modified-site, glycosylated or  
phosphorylated, wherein Xaa may be any  
amino acid residue, preferably Arg.  
(B) LOCATION: 16

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Lys Pro Gly Leu Asp Thr Tyr Ser Leu Gly Gly Gly Gly Ala Ala Xaa Val  
1 5 10 15

Arg

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 15 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(iii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(ix) FEATURE:

(A) NAME/KEY: Modified-site, glycosylated or  
phosphorylated, wherein Xaa may be any  
amino acid residue, preferably Ala

(B) LOCATION: 12

(ix) FEATURE:

(A) NAME/KEY: Modified-site, glycosylated or  
phosphorylated, wherein Xaa may be any  
amino acid residue, preferably Ser

(B) LOCATION: 14

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Trp	Arg	Leu	Leu	Leu	Thr	Ser	Ala	Pro	Ser	Leu	Xaa	Thr	Xaa	Pro
1					5					10				15